STM JOURNALS

Research & Reviews: Journal of
Statistics

https://journals.stmjournals.com/rrjost/

ISSN: 2278-2273 (Online)
ISSN: 2348-7909 (Print)
Volume 13, Issue 1, 2024
January—April
DOI (Journal): 10.37591/RRJoST

Research RRJoST

Size-biased Sujatha Distribution with Properties and Application to Model Flood Data

Rama Shanker[1], Hosenur Rahman Prodhani[2,*]

Abstract

In this study, a size-biased version of the Sujatha distribution was proposed to model flood data. The descriptive statistical properties based on moments and the reliability properties of the distribution are discussed in detail along with their derivation and graphical presentation. An interesting feature of the proposed distribution is that it is a member of the exponential family of distributions. A sequential probability ratio test was performed using the proposed distribution. There has been discussion of parameter estimation utilizing the moment, maximum likelihood, maximum product spacing, least squares, weighted least squares, and Cramer-Von Mises estimation methods. The confidence interval of the parameter of the distribution was obtained and shown graphically. A simulation study was conducted to determine the consistency of the estimator using different estimation methods has been done. To demonstrate the application of the distribution, we presented its goodness-of-fit by comparing it with that of the size-biased exponential and Lindley distributions. The results indicate that the proposed distribution provides the best fit among size-biased distributions. This study makes a significant contribution to the field by introducing a novel distribution model specifically designed for flood data. It offers an in-depth statistical analysis, explores various methods for estimating distribution parameters, and validates its practical use through simulations and goodness-of-fit assessments. The findings underscore the benefits of employing this new size-biased Sujatha distribution for modeling flood events and potentially other datasets in which size bias plays a crucial role.

Keywords: Sujatha distribution, mean residual life function, moments-based measures, sequential probability ratio test, estimation of parameter, applications

INTRODUCTION

The initial concept of weighted distribution was introduced by Fisher (1934) [1] to explain ascertainment bias and later on, it was Rao (1965) [2] who extended the idea of weighted distribution to model data in which classical standard distributions were not suitable for the observations with unequal probabilities of selection. To account for observations in such instances, weighted models were developed using a weight function. Biased data can arise from the frequency distribution of recorded data, such as at least one boy child per family, at least one girl child per family, at least one migration per family, or at least one heavy flood per year.

Suppose the original observation y is from a distribution having the probability density (pdf) $g_y(x,\eta_1)$, η_1 may, be a vector of parameters. Again, suppose the observation x is recorded based on the probability which is re-weighted by the weight function $w(x,\eta_2)>0$, η_2 being the new vector of parameters.

$$f(x;\eta_1,\eta_2)=Dw(x;\eta_2)g_y(x;\eta_1)$$

*Author for Correspondence
Hosenur Rahman Prodhani
E-mail: hosenur72@gmail.com

[1]Associate Professor, Department of Statistics, Assam University, Silchar, Assam, India
[2]Research Scholar, Department of Statistics, Assam University, Silchar, Assam, India

Received Date: July 18, 2024
Accepted Date: July 29, 2024
Published Date: August 05, 2024

Citation: Rama Shanker, Hosenur Rahman Prodhani. Size-biased Sujatha Distribution with Properties and Application to Model Flood Data. Research & Reviews: Journal of Statistics. 2024; 13(1): 31–46p.

Where, $D$is the normalizing vector. Such distributions are called weighted distributions. If we consider the weight function $w(x) = x$, then this is known as simple size-biased distribution.

The size-biased distribution becomes crucial when there is an underlying relationship between the size of the items and their chance of occurrence. A size-biased distribution offers a realistic representation of larger things that are more likely to be observed or chosen. For instance, in reliability engineering, larger machines or parts may have a higher failure rate, which is why reliability studies tend to find more failures. Larger fish may be more likely to be caught in nets while examining fish populations in ecology. In academic publishing, highly cited papers are more likely to be included in citation databases. In quality control, larger defects or failures in products might be more frequently detected and recorded. Size-biased distributions were applied by Van Duesen (1986) [3] to fit diameter at breast height data obtained from horizontal point sampling, and by Lappi and Bailey (1987) [4] to analyze diameter increment data obtained from the same horizontal point sampling. Patil and Rao (1977, 1978) [5, 6] extensively discussed the use of size-biased distributions in the interpretation of data pertaining to the human population and environment. Mir et al (2013) [7] introduced size-biased exponential distribution (SBED) and Ayesha (2017) [8] introduced size-biased Lindley distribution (SBLD).

Shanker (2016) [9] has introduced the Sujatha distribution and discussed its statistical properties and applications and concluded that the Sujatha distribution provides a better fit as compared to the Akash distribution by Shanker (2015) [10], Shanker distribution by Shanker (2015) [11], Lindley distribution by Lindley (1958) [12] and exponential distribution. The probability density function (pdf) and cumulative density function (cdf) of the Sujatha distribution with scale parameter η are given as

$$f(x;\eta) = \frac{\eta^3}{\eta^2+\eta+2}(1 + x + x^2)e^{-nx}; x > 0, \eta > 0$$

$$F(x;\eta) = 1 - \left[1 + \frac{\eta x(\eta x+\eta+2)}{\eta^2+\eta+2}\right] e^{-nx}; x > 0, \eta > 0$$

Recently, several generalizations and modifications of Sujatha distribution have been derived and studied by several researchers including quasi-Sujatha distribution (QSD) by Shanker (2016) [13], a generalization of Sujatha distribution(GSD) by Shanker et al (2017), a two-parameter Sujatha distribution (TPSD) by Tesfay and Shanker (2018) [14], a new two-parameter Sujatha distribution (NTPSD) by Tesfay and Shanker (2018) [15], another new two-parameter Sujatha distribution (ANTPSD) by Shanker (2019) [16], weighted Sujatha distribution (WSD) by Shanker and Shukla (2018) [17], Power Sujatha distribution(PSD) by Shanker and Shukla (2019) [18], a new quasi-Sujatha distribution (NQSD) by Shanker and Shukla (2020) [19], generalized Inverse power Sujatha distribution(GIPSD) by Okoli et al (2021) [20], Marshal-Olkin Sujatha distribution (MOSD) by Ikechukwu and Eghwerido (2022) [21], exponentiated Sujatha distribution(ESD) by Prodhani and Shanker (2022) [22], are some among others.

The rationale behind proposing the size-biased Sujatha distribution (SBSD) is that a size-biased distribution allows for more accurate modeling of phenomena where the probability of selection is proportional to size or magnitude. This approach addresses the inherent sampling biases and provides more representative estimates and predictions. Some important statistical and mathematical properties of SBSD have been discussed along with an estimation of parameters by different methods and applications to model flood data.

SIZE-BIASED SUJATHA DISTRIBUTION

Following the approach of deriving the pdf of size-biased distribution, the pdf and the cdf of SBSD can be obtained as

$$f(x;\eta) = \frac{\eta^4}{\eta^2+2\eta+6}(1 + x + x^2)xe^{-\eta x}; x > 0, \eta > 0$$

$$F(x;\eta) = 1 - \left[1 + \frac{\eta x(\eta^2 x^2 + \eta^2 x + 3\eta x + 2\eta + 6)}{\eta^2 + 2\eta + 6}\right] e^{-\eta x}$$

The graphical representation of the pdf and cdf of the SBSD are shown in Figure 1. It is clear from the Figure 1 that for different values of the parameter η, the SBSD has unimodal, bimodal, and positively skewed natures.

RELIABILITY PROPERTIES

Reliability Function

The reliability function of SBSD and its graph can be presented in Figure 2.

$$R(x;\eta) = \left[1 + \frac{\eta x(\eta^2 x^2 + \eta^2 x + 3\eta x + 2\eta + 6)}{\eta^2 + 2\eta + 6}\right] e^{-\eta x}; x > 0, \eta > 0$$

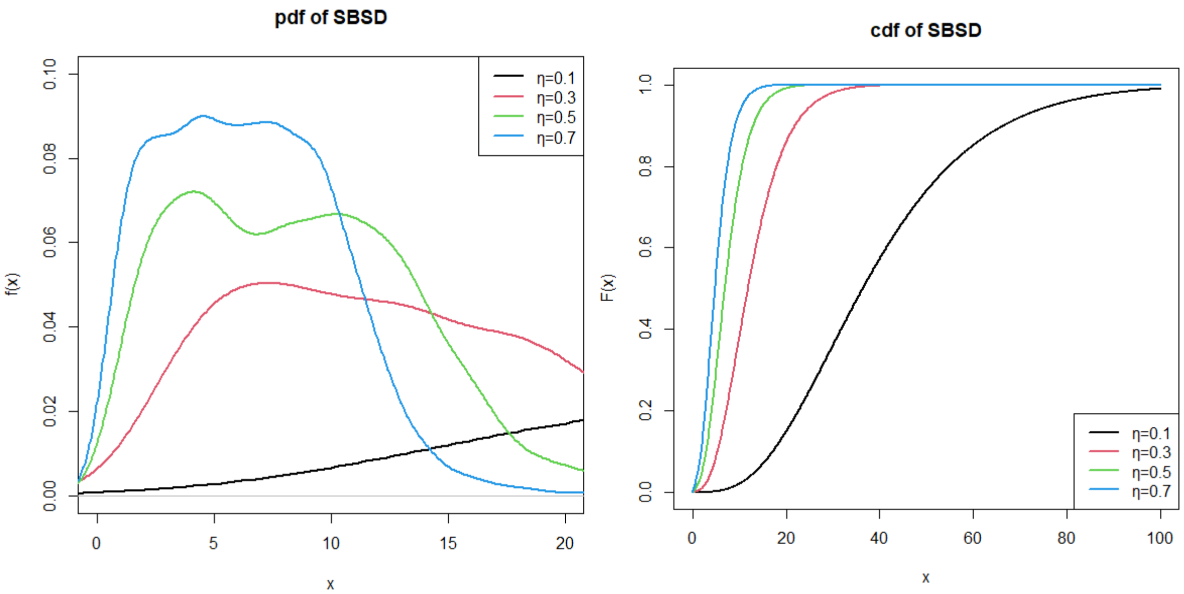


Figure 1. pdf and cdf of SBSD.

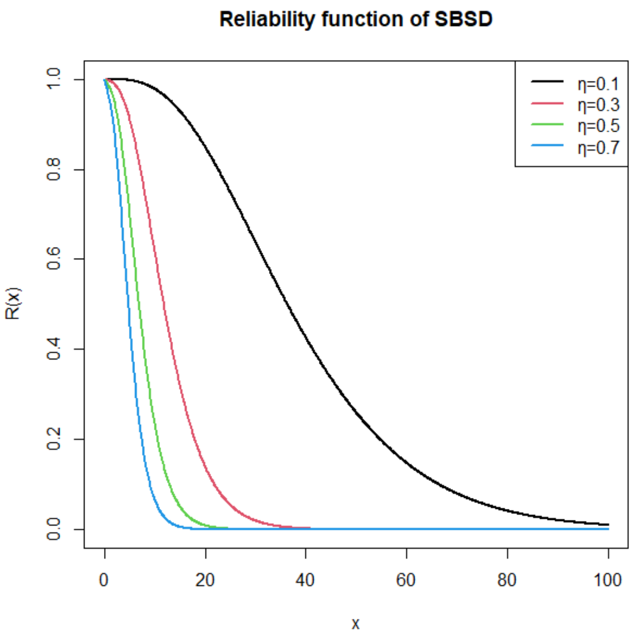


Figure 2. Reliability function of SBSD.

Hazard Function

The hazard function of SBSD can be obtained as

$$h(x;\eta) = \frac{\eta^4(1+x+x^2)x}{\eta^2+2\eta+6+\eta x(\eta^2x^2+\eta^2x+3\eta x+2\eta+6)}; x > 0, \eta > 0$$

It is clear from Figure 3 that for increasing values of the parameter η the hazard function is strictly increasing.

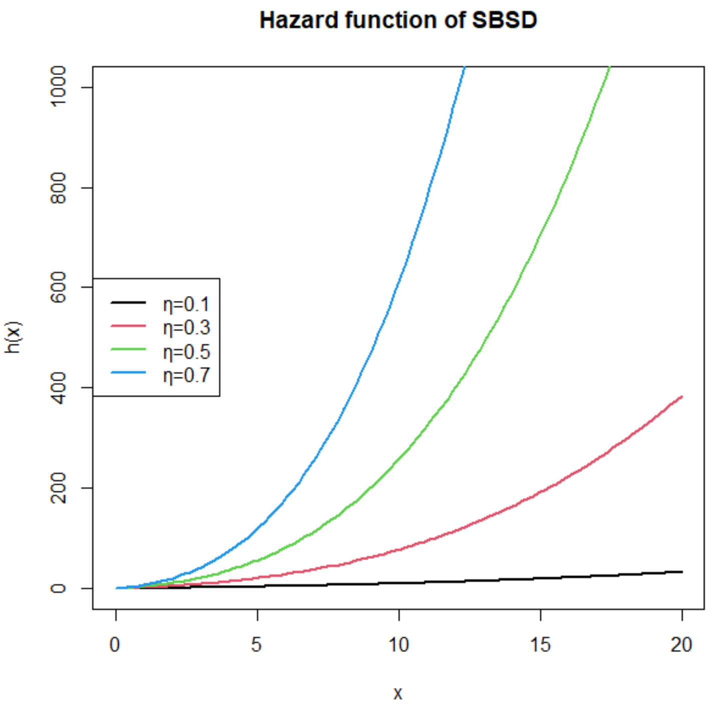


Figure 3. Hazard function of SBSD.

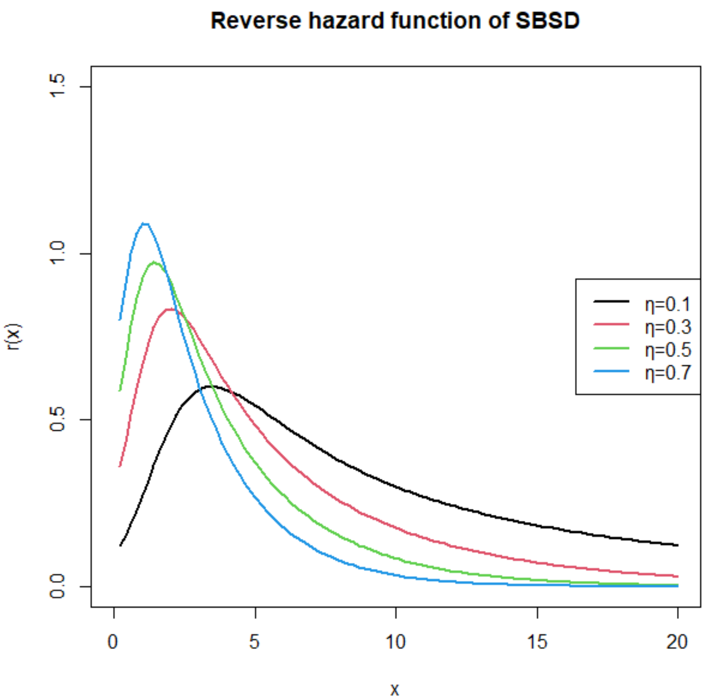


Figure 4. Reverse hazard function of SBSD.

Reverse Hazard Function

The reverse hazard function of SBSD can be obtained as

$$h(x;\eta) = \frac{\eta^4(1+x+x^2)xe^{-\eta x}}{\eta^2+2\eta+6-[\eta^2+2\eta+6+\eta x(\eta^2x^2+\eta^2x+3\eta x+2\eta+6)]e^{-\eta x}}; x>0, \eta>0$$

The graphical representation of the reverse hazard function is presented in Figure 4.

It is clear from Figure 4 that for increasing values of the parameter η the reverse hazard function first increases after that decreases.

Mean Residual Life Function

The mean residual life function of SBSD can be obtained as

$$m(x;\eta) = \frac{1}{1-F(x;\eta)}\int_t^{\theta} tf(t;\eta)\,dt - x$$

$$= \frac{(x^4+x^3+x^2)\eta^4+(4x^3+3x^2+2x)\eta^3+2(6x^2+3x+x)\eta^2+6(4x+1)\eta+24}{\eta[\eta^2+2\eta+6+\eta x(\eta^2x^2+\eta^2x+3\eta x+2\eta+6)]} - x$$

For $m(0,\eta) = \frac{2\eta^2+6\eta+24}{\eta(\eta^2+2\eta+6)}$ which is a mean of SBSD. The graphical representation of the mean residual life function of SBSD is shown in Figure 5. It is clear from Figure 5, mean residual life function is monotonically decreasing.

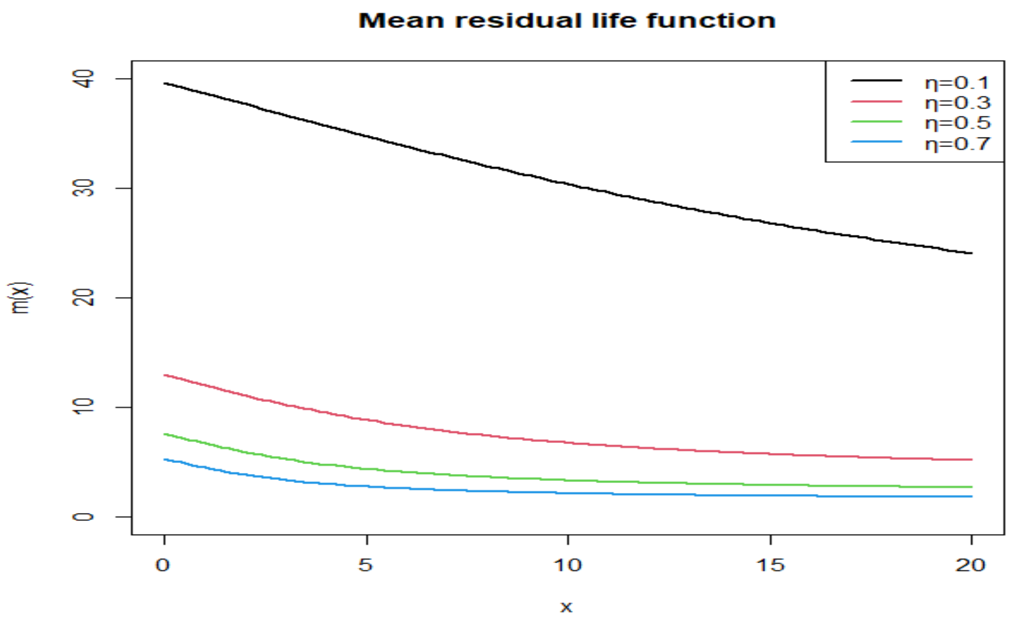


Figure 5. Mean residual life function of SBSD.

MOMENTS-BASED MEASURES

As we know the moments are essential for a distribution to determine the descriptive nature of the distribution including coefficient of variation, skewness, kurtosis, and index of dispersion. The rth moment about the origin $\mu_r{}'$ of SBSD can be obtained as

$$\mu_r{}' = E(X^r) = \frac{\eta^4}{\eta^2+2\eta+6}\int_0^{\infty} x^r(1+x+x^2)\,xe^{-\theta x}dx$$

$$= \frac{(r+1)!\{\eta^2+(r+2)\eta+(r+2)(r+3)\}}{\theta^r(\eta^2+2\eta+6)}; r = 1,2,3,\cdots$$

Substituting $r = 1,2,3,4$ in the above expression, the first four moments about the origin of SBSD can be obtained as

$$\mu_1{}' = \frac{2\eta^2+6\eta+24}{\eta(\eta^2+2\eta+6)},\ \mu_2{}' = \frac{6\eta^2+24\eta+120}{\eta^2(\eta^2+2\eta+6)}$$

$$\mu_3{}' = \frac{24\eta^2+120\eta+720}{\eta^3(\eta^2+2\eta+6)},\ \mu_4{}' = \frac{120\eta^2+720\eta+5040}{\eta^4(\eta^2+2\eta+6)}.$$

The moments about the mean of SBSD can thus be obtained as

$$\mu_2 = \frac{2(\eta^4+6\eta^3+36\eta^2+48\eta+72)}{\eta^2(\eta^2+2\eta+6)^2}$$

$$\mu_3 = \frac{4(\eta^6+9\eta^5+72\eta^4+174\eta^3+360\eta^2+432\eta+432)}{\eta^3(\eta^2+2\eta+6)^3}$$

$$\mu_4 = \frac{24(\eta^8+127\eta^7+114\eta^6+468\eta^5+1530\eta^4+3216\eta^3+5472\eta^2+5184\eta+3888)}{\eta^4(\eta^2+2\eta+6)^4}$$

Thus, the coefficient of variation (CV), coefficient of skewness $\left(\sqrt{\beta_1}\right)$, coefficient of kurtosis(β_2), and index of dispersion (γ) of SBSD is thus obtained as

$$CV = \frac{\sqrt{2(\eta^4+6\eta^3+36\eta^2+48\eta+72)}}{2\eta^2+6\eta+24}$$

$$\sqrt{\beta_1} = \frac{4(\eta^6+9\eta^5+72\eta^4+174\eta^3+360\eta^2+432\eta+432)}{[2(\eta^4+6\eta^3+36\eta^2+48\eta+72)]^{3/2}}$$

$$\beta_2 = \frac{24(\eta^8+127\eta^7+114\eta^6+468\eta^5+1530\eta^4+3216\eta^3+5472\eta^2+5184\eta+3888)}{[2(\eta^4+6\eta^3+36\eta^2+48\eta+72)]^2}$$

$$\gamma = \frac{2(\eta^4+6\eta^3+36\eta^2+48\eta+72)}{\eta(\eta^2+2\eta+6)(2\eta^2+6\eta+24)}.$$

It has been observed that SBSD is under-dispersed $(\mu > \sigma^2)$, equi-dispersed $(\mu = \sigma^2)$ and over-dispersed $(\mu < \sigma^2)$ for $\eta(>,=,<)0.8546$. The behaviors of the coefficient of variation, coefficient of skewness, coefficient of kurtosis, and index of dispersion for SBSD are presented in Figure 6. As the values increase, the coefficient of variation increases, while the coefficient of skewness and coefficient of kurtosis initially decrease and then increase in a U-shape pattern. The index of dispersion decreases.

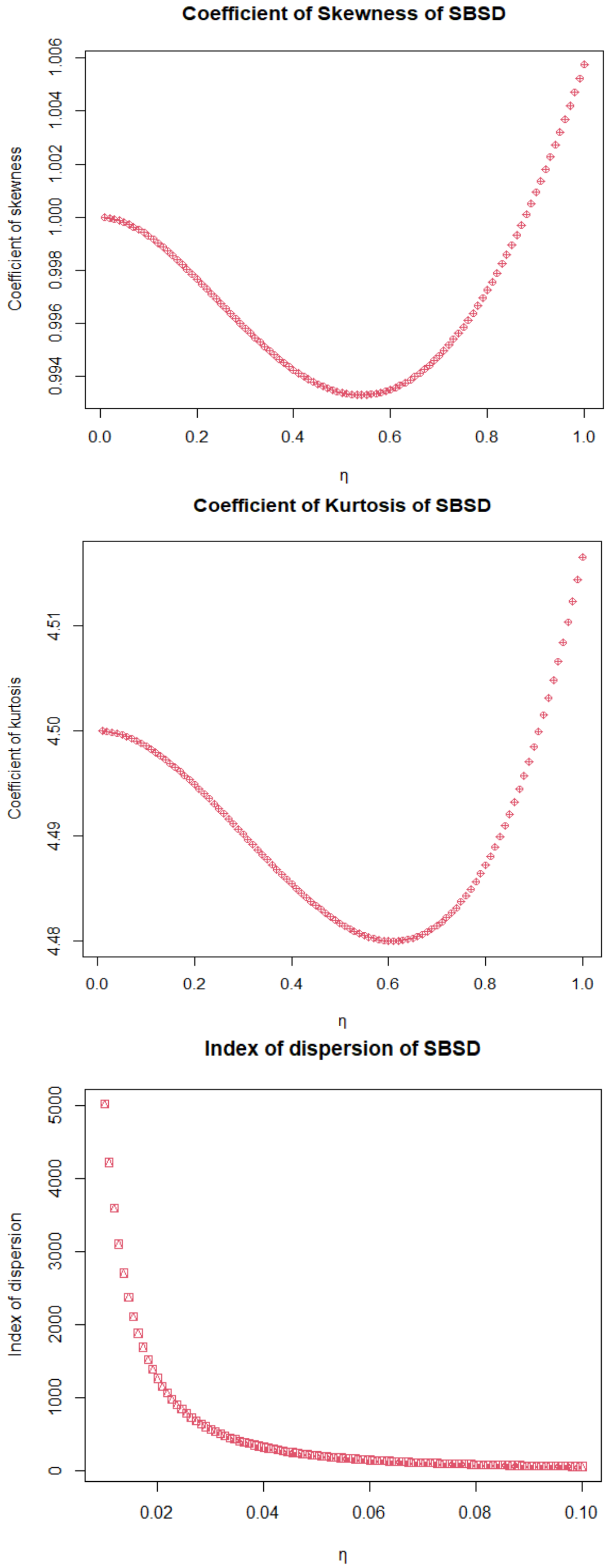


Figure 6. Coefficient of variation, skewness, kurtosis, and index of dispersion of SBSD.

STATISTICAL INFERENCE

Exponential Family of Distribution

We can show that SBSD is a member of the exponential family of distribution. We have

$$f(x;\eta) = \frac{\eta^4}{\eta^2+2\eta+6}(1+x+x^2)xe^{-\eta x} = a(\eta)b(x)e^{-c(\eta)d(x)}$$

Where, $a(\eta) = \frac{\eta^4}{\eta^2+2\eta+6}$, $b(x) = (1+x+x^2)x$, $c(\eta) = \eta$, $d(x) = x$.

Thus, SBSD belongs to an exponential family, and hence, $\sum_{i=1}^{n} d(x_i) = \sum_{i=1}^{n} x_i$ is sufficient, complete, and minimal sufficient statistic.

Sequential Probability Ratio Test

Let X_i be a random variable that follow SBSD (η). Then the joint pdf of SBSD can be expressed as

$$\prod_{i=1}^{n} f(x_i;\eta) = \left(\frac{\eta^4}{\eta^2+2\eta+6}\right)^n \prod_{i=1}^{n}(1+x_i+{x_i}^2)x_i\, e^{-\eta\sum_{i=1}^{n} x_i}$$

Now,

$$L_m = \frac{\prod_{i=1}^{n} f(x_i;\eta_1)}{\prod_{i=1}^{n} f(x_i;\eta_0)} = \left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)^n e^{(\eta_0-\eta_1)\sum_{i=1}^{n} x_i}$$

We have to reject H_0 if.

$$L_m = \left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)^n e^{(\eta_0-\eta_1)\sum_{i=1}^{n} x_i} \geq A$$

$$(\eta_0-\eta_1)\sum_{i=1}^{n} x_i \geq \log A - n\log\left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)$$

If $\eta_1 > \eta_0$, then $\eta_0 - \eta_1 < 0$, so for this case, we have to reject H_0 if.

$$\sum_{i=1}^{n} x_i \leq \frac{\log A - n\log\left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)}{(\eta_0-\eta_1)}$$

If $\eta_1 < \eta_0$, then $\eta_0 - \eta_1 > 0$, so for this case, we have to reject H_0 if.

$$\sum_{i=1}^{n} x_i \geq \frac{\log A - n\log\left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)}{(\eta_0-\eta_1)}$$

We have to accept H_0 if.

$$L_m = \left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)^n e^{(\eta_0-\eta_1)\sum_{i=1}^{n} x_i} \leq B$$

$$(\eta_0-\eta_1)\sum_{i=1}^{n} x_i \leq \log B - n\log\left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)$$

If $\eta_1 > \eta_0$, then $\eta_0 - \eta_1 < 0$, so for this case, we have to accept H_0 if.

$$\sum_{i=1}^{n} x_i \geq \frac{\log B - n\log\left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)}{(\eta_0-\eta_1)}$$

and if $\eta_1 < \eta_0$, then $\eta_0 - \eta_1 > 0$, so for this case, we have to accept H_0 if.

$$\sum_{i=1}^{n} x_i \leq \frac{\log B - n\log\left(\frac{{\eta_1}^4({\eta_0}^2+2\eta_0+6)}{{\eta_0}^4({\eta_1}^2+2\eta_1+6)}\right)}{(\eta_0-\eta_1)}.$$

Here the constants A and B are determined based on the probability of type I error and the probability of type II error.

ESTIMATION OF THE PARAMETER

In this section following methods are used to estimate the parameter of the SBSD

Maximum Likelihood Estimation

Let $(x_1, x_2, \dots, x_n)$be a random sample from SBSD(η). The log-likelihood function of SBSD can be obtained as

$$log\, L = 4n\, log\, \theta - n\, log(\eta^2 + 2\eta + 6) + \sum_{i=1}^{n} log(1 + x + x^2) + \sum_{i=1}^{n} log\, x - \eta \sum_{i=1}^{n} x$$

Now, $\frac{d}{d\eta} log\, L = \frac{4n}{\eta} - \frac{2n(\eta+1)}{\eta^2+2\eta+6} - n\bar{x}$

The maximum likelihood estimates (MLE), $\hat{\eta}$of $\eta$is the solution of the equation $\frac{d}{d\eta} log\, L = 0$and is given by the solution of the following cubic equation.

$$\bar{x}\eta^3 + 2(\bar{x} - 1)\eta^2 + 6(\bar{x} - 1)\eta - 24 = 0$$

The one important characteristic of SBSD is that the method of moment estimates (MOME) $\hat{\eta}$ of η is the same as given by MLE.

Maximum Product Spacing Estimation

The maximum product spacing estimates (MPSE) $\hat{\eta}$ of η can be obtained numerically by maximizing the following function in relation to η

$$MPS = \frac{1}{n+1} \sum_{i=1}^{n+1} log[F(x_i, \eta) - F(x_{i-1}, \eta)]$$

Least Squares Estimation

Let $(x_1, x_2, \dots, x_n)$be a random sample from SBSD(η). The least square function of the parameter based on the pdf of SBSD can be given as

$$LSE = \sum_{i}^{n} \left(F\left(x_{(i)}\right) - \frac{i}{n+1} \right)^2$$

The least square estimates (LSE) $\hat{\eta}$ of η can be obtained by minimizing the above equation with respect to η.

Weighted Least Squares Estimation

Let $(x_1, x_2, \dots, x_n)$be a random sample from SBSD(η). The weighted least square function of the parameter based on the pdf of SBSD can be given as

$$WLSE = \sum_{i}^{n} w_i \left(F\left(x_{(i)}\right) - \frac{i}{n+1} \right)^2, \text{ where } \quad w_i = \frac{(n+1)^2(n+2)}{i(n-i+1)}$$

The weighted least square estimates (WLSE) $\hat{\eta}$ of η can be obtained by minimizing the above equation with respect to η.

Cramer-Von Mises Estimation

Let $(x_1, x_2, \dots, x_n)$be a random sample from SBSD(η). The Cramer-Von Mises function of the parameter based on the pdf of SBSD can be given as

$$CVM = \frac{1}{12n} + \sum_{i=1}^{n} F\left(x_{(i)} - \frac{2i-1}{2n} \right)$$

The Cramer-Von Mises estimates (CVME) $\hat{\eta}$ of η can be obtained by minimizing the above equation with respect to η.

THE SIMULATION STUDY

The simulation study based on acceptance-rejection method has been carried out to know the consistency of parameter estimator obtained by different methods of estimation. It is based on examining the biases (B) and the mean squares (MSE) of different estimator of the parameter of SBSD using the following formulas.

$$B = \frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}-\eta), \text{ MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}-\eta)^2$$

The following procedures were used to create the random samples from the SBSD.

- Generate Y from exponential (η) distribution.
- Generates U from Uniform (0,1) distribution.
- If $U \le \frac{f(y)}{M\,g(y)}$, then set $X = Y$ ("accept the sample"); otherwise ("reject the sample") and if reject then repeat the process: step (a-c) until getting the required samples. Where M is a constant.
- Each sample size is replicated 10000 times

The bias and the MSE of the parameter of SBSD, obtained from MLE, MPSE, LSE, WLSE and CVME, are decreasing for increasing sample size, which is obvious from Table 1.

From the Table 1, it observed that when the parameter $\eta = 0.1$**,** LSE, MLE, MPSE provides better estimate as compared WLSE, CVME, when the parameter $\eta = 0.5$**,** MPSE provides better estimate as compared to MLE, LSE, WLSE and CVME, when the parameter $\eta = 1.0$**,** CVME provides better estimate as compared to MLE, MPSE, LSE and WLSE.

Table 1. Descriptive constants of the parameters of SBSD.

η	n	MLE		MPSE		LSE		WLSE		CVME	
		Bias	***MSE***	***Bias***	***MSE***	***Bias***	***MSE***	***Bias***	***MSE***	***Bias***	***MSE***
0.1	20	-0.01928	0.00040	-0.01929	0.00040	-0.01903	0.00039	-0.02328	0.00060	-0.02324	0.00057
	40	-0.01898	0.00038	-0.01900	0.00038	-0.01875	0.00037	-0.02218	0.00055	-0.02216	0.00052
	60	-0.01848	0.00036	-0.01871	0.00037	-0.01850	0.00036	-0.02135	0.00051	-0.02078	0.00047
	80	-0.01798	0.00035	-0.01769	0.00034	-0.01765	0.00034	-0.02043	0.00048	-0.01995	0.00043
	100	-0.01626	0.00030	-0.01662	0.00032	-0.01618	0.00030	-0.01953	0.00044	-0.01926	0.00040
0.5	20	0.02479	0.00064	0.02065	0.00047	0.03387	0.00118	-0.01326	0.00141	-0.04963	0.00251
	40	0.02399	0.00060	0.01953	0.00042	0.03269	0.00109	-0.01103	0.00127	-0.04749	0.00231
	60	0.02327	0.00056	0.01853	0.00038	0.03156	0.00102	-0.00832	0.00123	-0.04595	0.00217
	80	0.02235	0.00052	0.01774	0.00035	0.03026	0.00094	-0.00781	0.00110	-0.04435	0.00202
	100	0.02182	0.00049	0.01709	0.00032	0.02932	0.00089	-0.00740	0.00100	-0.04321	0.00192
1.0	20	0.14080	0.02117	0.09841	0.01064	0.15878	0.02633	0.06883	0.02485	0.01599	0.00049
	40	0.13584	0.01987	0.09168	0.00971	0.14736	0.02331	0.06655	0.01811	0.01411	0.00047
	60	0.12812	0.01794	0.08960	0.00935	0.14367	0.02231	0.05891	0.01531	0.01273	0.00041
	80	0.12328	0.01676	0.08647	0.00881	0.13773	0.02070	0.05748	0.01381	0.01127	0.00037
	100	0.11495	0.01475	0.07957	0.00772	0.13014	0.01883	0.05455	0.01285	0.00961	0.00035

A REAL APPLICATION

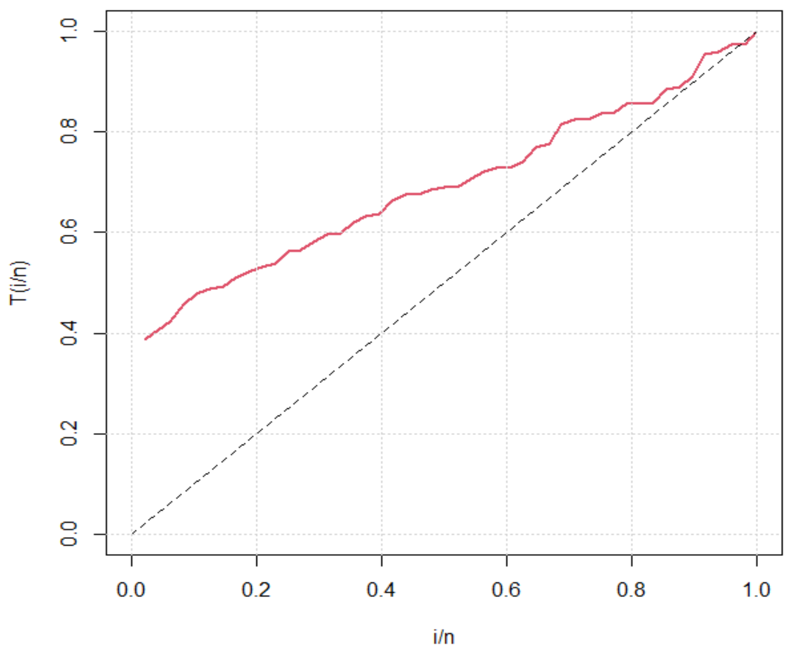
To demonstrate how SBSD is applied to flood data, we have used an example. The following right-skewed data set, which Van Montfort (1970) [23] discussed, shows the North Saskatchewan River's greatest annual flood flows over a 47-year period near Edmonton, measured in cubic feet per second. The data is:

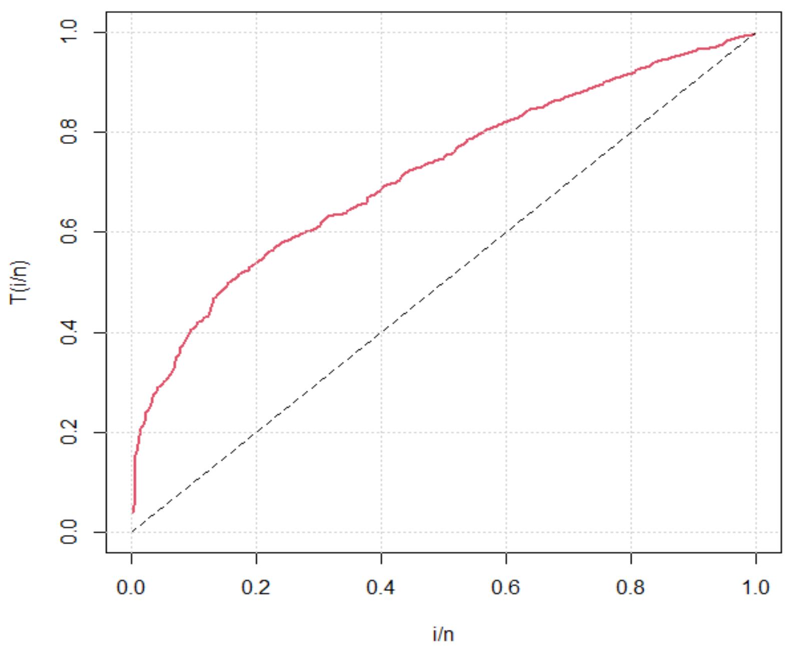
Research & Reviews: Journal of Statistics
Volume 13, Issue 1
ISSN: 2278-2273 (Online), ISSN: 2348-7909 (Print)

19.885, 20.940, 21.820, 23.700, 24.888, 25.460, 25.760, 26.720, 27.500, 28.100, 28.600, 30.200, 30.380, 31.500, 32.600, 32.680, 34.400, 35.347, 35.700, 38.100, 39.020, 39.200, 40.000, 40.400, 40.400, 42.250, 44.020, 44.730, 44.900, 46.300, 50.330, 51.442, 57.220, 58.700, 58.800, 61.200, 61.740, 65.440, 65.597, 66.000, 74.100, 75.800, 84.100, 106.600, 109.700, 121.970, 121.970, 185.560.

The total time on test (TTT) plots and the histogram of the original dataset and the corresponding simulated dataset are shown in Figure 7.

A TTT- plot that curves upward suggests that the failure rate is decreasing over time, indicative of early failures or "infant mortality". This implies that the system is improving with time as weak units fail early. From Figure 7, the sample dataset and theoretical values both have decreasing failure rate.





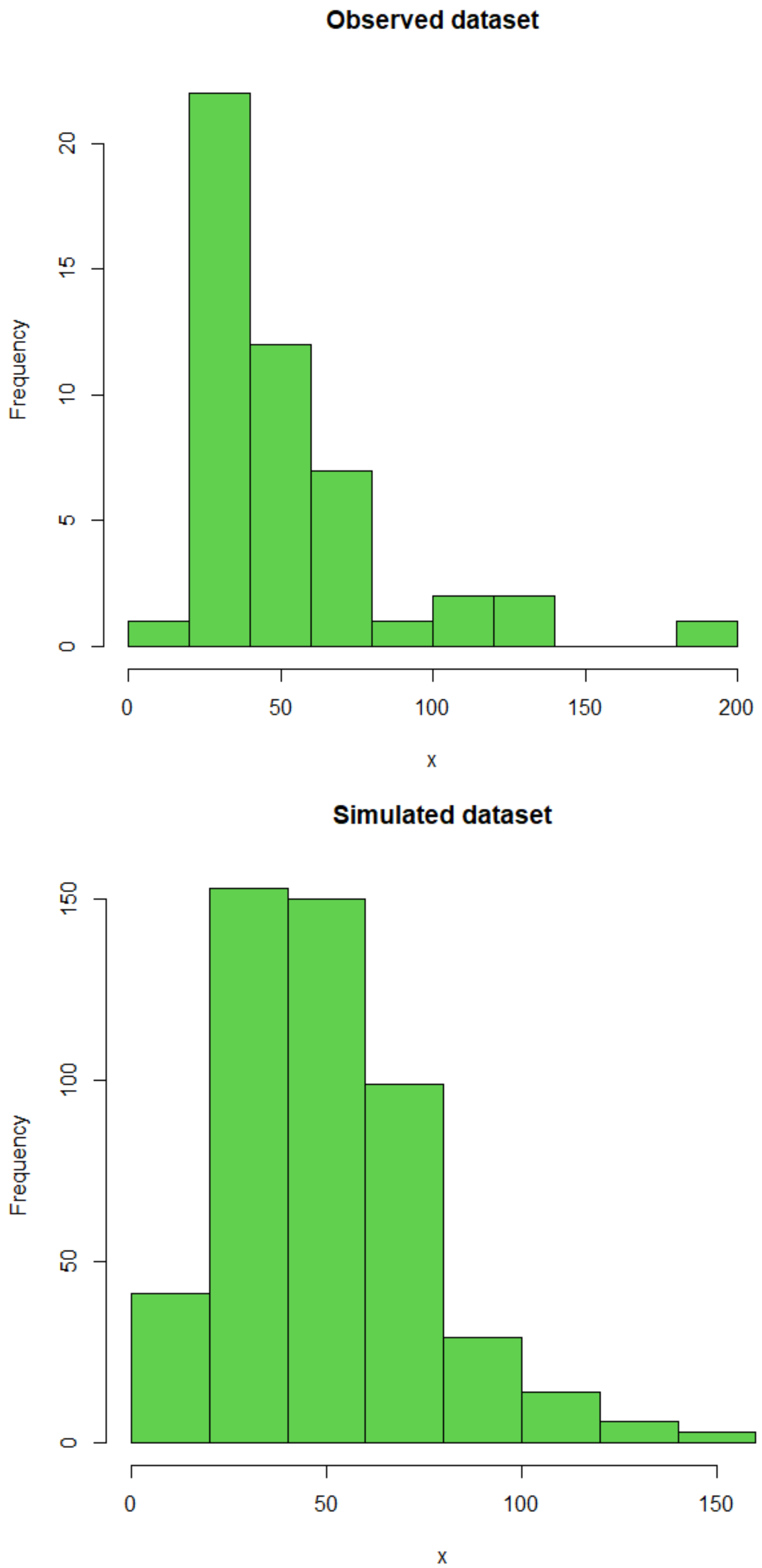


Figure 7. TTT-plot and histogram of the observed and theoretical values of the dataset.

We calculated the values of $-2\log l$ the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), Consistent Akaike Information Criterion (CAIC), Kolmogorov-Smirnov Statistics (K-S), and the corresponding (p-value) for the dataset to compare lifetime distributions. The formulas used to calculate AIC, BIC, CAIC, HQIC, and K-S are as follows:

$$AIC = -2\,log\,L + 2p,\,BIC = -2\,log\,L + p\,log(n),\,CAIC = -2\,log\,L + \frac{2pn}{n-p-1}HQIC$$
$$= -2\,log\,L + 2p\,log[log(n)],K-S = \underset{x}{Sup}|F_m(x) - F_o(x)|$$

Where p is the number of parameters, p is the sample size, $F_m(x)$ is the empirical CDF, and $F_p(x)$ is the CDF of the considered distribution. Table 2 presents the estimates provided by different estimation techniques along with their standard errors (in parentheses) for the distributions. The goodness-of-fit measures for the dataset are shown in Table 3.

Table 3 clearly shows that SBSD has the lowest -2 log l AIC, BIC, CAIC, HQIC, and K-S values among the considered distributions, indicating that SBSD provides the best fit.

Table 2. MLE, MPSE, LSE, WLSE, CME of parameter with their standard errors (in parenthesis) of the parameter of the considered distribution of the dataset.

Distributions	MLE $\hat{\eta}$ $SE(\hat{\eta})$	MSPE $\hat{\eta}$ $SE(\hat{\eta})$	LSE $\hat{\eta}$ $SE(\hat{\eta})$	WLSE $\hat{\eta}$ $SE(\hat{\eta})$	CME $\hat{\eta}$ $SE(\hat{\eta})$
SBSD	0.0771 (0.0055)	0.0758 (0.0054)	0.0826 (0.0022)	0.0817 (0.0008)	0.0843 (0.0146)
SBLD	0.0577 (0.0048)	0.0691 (0.0011)	0.0688 (0.0016)	0.0693 (0.0006)	0.0695 (0.0109)
SBED	0.0388 (0.0039)	0.1210 (0.0010)	0.1200 (0.0010)	0.1425 (0.0003)	0.1300 (0.0074)
Sujatha	0.0576 (0.0048)	0.0691 (0.0011)	0.0593 (0.0017)	0.0579 (0.0006)	0.0607 (0.0115)
Lindley	0.0381 (0.0038)	0.0372 (0.0037)	0.0364 (0.0012)	0.0349 (0.0004)	0.0373 (0.0081)
Exponential	0.0194 (0.0027)	0.0187 (0.0027)	0.0151 (0.0007)	0.0145 (0.0002)	0.0158 (0.0047)

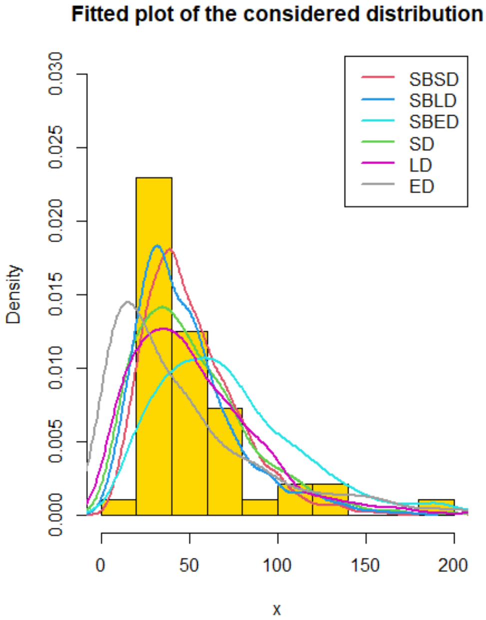
Table 3. Goodness-of-fit of the dataset.

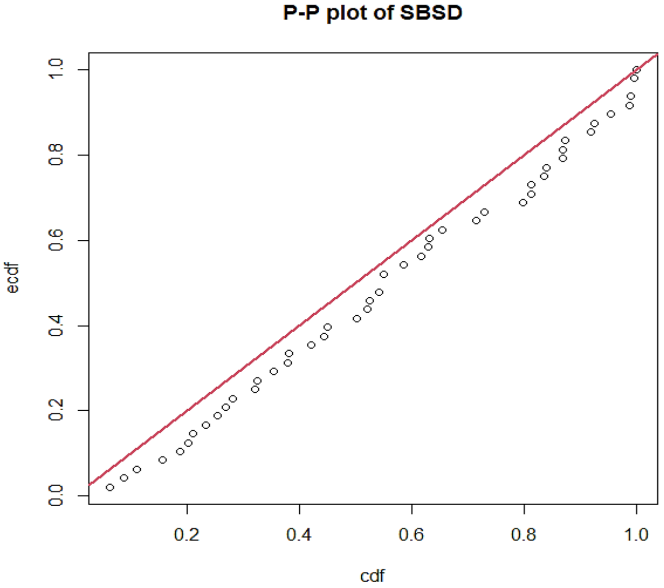
Distributions	$-2\,log\,L$	AIC	BIC	CAIC	HQIC	K-S	p-value
SBSD	443.34	445.34	452.44	445.42	446.04	0.12	0.45
SBLD	444.35	446.35	453.45	446.43	447.05	0.13	0.38
SBED	451.03	453.03	460.13	453.11	453.73	0.33	0.00
Sujatha	444.44	446.44	453.54	448.52	449.14	0.14	0.31
Lindley	452.28	454.28	461.38	454.36	454.98	0.19	0.07
Exponential	474.38	476.38	483.48	476.46	477.08	0.31	0.00

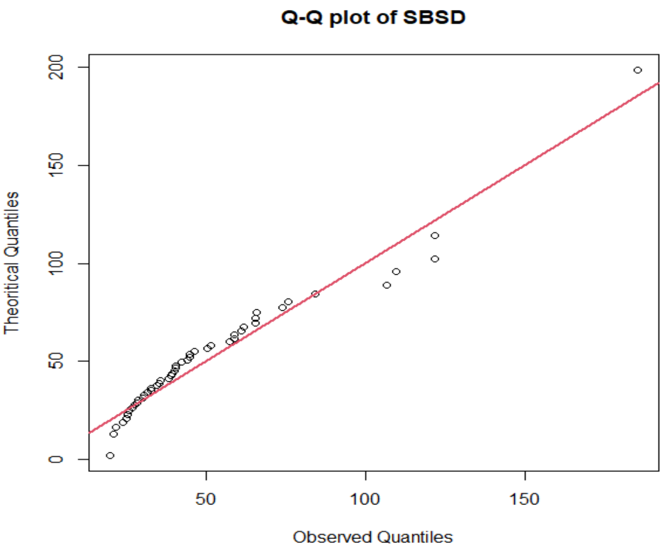
Table 4. Confidence interval of the parameters of SBSD for the dataset

Parameters	90% CI (Lower, Upper)	95% CI (Lower, Upper)	99% CI (Lower, Upper)
$\hat{\eta}$	0.0683, 0.0866	0.0667, 0.0885	0.0636, 0.0923

Figure 8 displays the fitted plot, P-P plot, Q-Q plot, and ECDF plot of the dataset, all of which support the hypothesis that the SBSD offers the best fit among the considered distributions. The confidence interval for the SBSD parameter is provided in Table 4. The profile plot of the parameter of SBSD for the dataset is given in Figure 9.


Fitted plot of the considered distribution
Density
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0
50
100
150
200
x
SBSD
SBLD
SBED
SD
LD
ED


P-P plot of SBSD
ecdf
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.4
0.6
0.8
1.0
cdf


Q-Q plot of SBSD
Theoritical Quantiles
0
50
100
150
200
50
100
150
Observed Quantiles

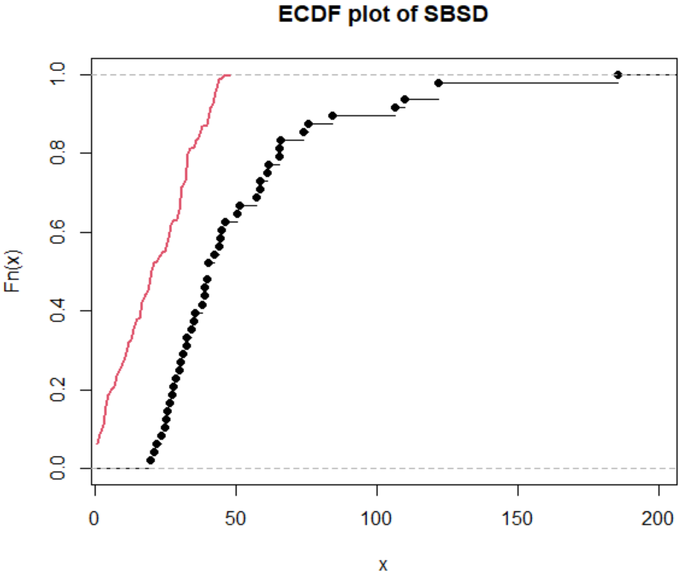


Figure 8. Fitted plot of the considered distributions, P-P plot, Q-Q plot and ECDF plot of the dataset.

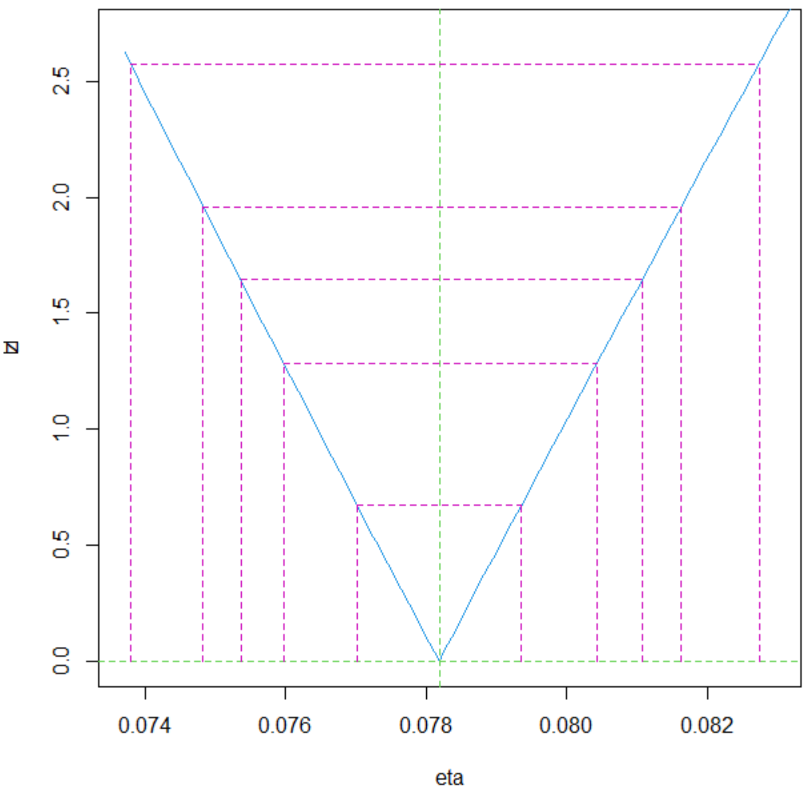


Figure 9. Profile plot of the dataset.

CONCLUSION

In this paper, the SBSD has been proposed to model flood data. Studies have been conducted on the mean residual life function, hazard function, reverse hazard function, reliability function, and moment-related metrics. The SBSD belongs to the exponential family of distribution. The sequential probability ratio test of SBSD is also discussed. Parameters are estimated by various methods including maximum likelihood estimation, maximum product spacing estimation, least square estimation, weighted least square estimation, and Cramer-Von Mises estimation. The confidence interval of the parameter is given along with its profile plot. The simulation study shows the consistency of the estimator. Based on the distribution's goodness-of-fit analysis using actual flood data, the SBSD distribution is more suitable than the SBLD, SBED, Sujatha, Lindley, and exponential distributions. Therefore, SBSD can be considered an important probability model for flood data.

REFERENCES

1. Fisher RA. The effect of methods of ascertainment upon the estimation of frequencies. Ann Eugen. 1934;6(1):13–25.
2. Rao CR. On discrete distributions arising out of methods of ascertainment. Sankhya A. 1965;27(4):311–24.
3. Van Deusen PC. Fitting assumed distributions to horizontal point sample diameters. For Sci. 1986;32(1):146–8.
4. Lappi J, Bailey RL. Estimation of the diameter increment function or other tree relations using angle-count samples. For Sci. 1987;33(3):725–39.
5. Patil GP. The weighted distribution: A survey of their applications. Biometrics. 1976;32(3):179–89.
6. Patil GP, Rao CR. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics. 1978;34(2):179–89.
7. Mir KA, Ahmed A, Reshi JA. On size-biased exponential distribution. J Mod Math Stat. 2013;7(2):21–5.
8. Ayesha A. Size biased Lindley distribution and its properties: A special case of weighted distribution. Appl Math. 2017;8(6):808–19.
9. Shanker R. Sujatha distribution and its applications. Stat Transit New Ser. 2016;17(3):391–410.
10. Shanker R. Shambhu distribution and its applications. Int J Probab Stat. 2016;5(2):48–63.
11. Shanker R. Shanker distribution and its applications. Int J Stat Appl. 2015;5(6):338–48.
12. Lindley DV. Fiducial distributions and Bayes' theorem. J R Stat Soc Ser B Methodol. 1958;20(1):102–7.
13. Shanker R. A quasi Sujatha distribution. Int J Probab Stat. 2016;5(4):89–100.
14. Tesfay M, Shanker R. A two-parameter Sujatha distribution. Biom Biostat Int J. 2018;7(4):188–97.
15. Tesfay M, Shanker R. A new two-parameter Sujatha distribution. Turkiye Klinikleri J Biostat. 2018;10(2):96–113.
16. Shanker R, Shukla KK. A two-parameter power Sujatha distribution with properties and application. Int J Math Stat. 2019;20(3):11–22.
17. Shanker R, Shukla KK. A two-parameter weighted Sujatha distribution and its application. Int J Appl Math Stat. 2018;57(3):106–21.
18. Shanker R, Tesfay M. Another two-parameter Sujatha distribution with properties and applications. J Math Sci Model. 2019;2(1):1–3.
19. Shanker R. A new quasi Sujatha distribution. Stat Transit New Ser. 2020;21(3):53–71.
20. Okoli OM, Osuji GA, Onyekwere CK. Generalized inverse power Sujatha distribution with applications. Asian J Probab Stat. 2021;15(3):11-25. doi: 10.9734/ajpas/2021/v15i330354.
21. Ikechukwu AF, Eghwerido JT. Marshall-Olkin Sujatha distribution and its applications. Thail Stat. 2022;20(1):36–52.
22. Prodhani HR, Shanker R. Exponentiated Sujatha distribution and its applications. In: Proceedings of the Eighth International Conference on Statistics for Twenty-First Century; 2022; Trivandrum, Kerala, India.
23. Van Montfort MA. On testing that the distribution of extremes is of type I when type II is the alternative. J Hydrol. 1970;11(4):421–7.

